CORRELATION ESTIMATION FOR DISTRIBUTED SOURCE CODING UNDER INFORMATION EXCHANGE CONSTRAINTS

Ngai-Man Cheung, Huisheng Wang and Antonio Ortega

Dept. of Electrical Engineering, Univ. of Southern California, Los Angeles, CA

ABSTRACT

Distributed source coding (DSC) depends strongly on accurate knowledge of correlation between sources. Previous works have reported capacity-approaching code constructions when *exact* knowledge of correlation is available at the encoder. However, in many applications exact correlation information may not be available, and correlation estimation is necessary. While error in estimation is inevitable, the impact of estimation error on compression efficiency has not been sufficiently studied for the DSC problem. In this paper we study correlation estimation subject to complexity constraints, and its impact on coding efficiency in a DSC framework. In particular, we consider the case where estimation entails information exchange between spatially separate sources and thus correlation estimation is subject to rate constraints. We first derive optimal strategies for information exchange that minimize the rate penalty due to inaccurate estimation, under constraints on the number of bits that can be exchanged between sources. Experimental results show that significant gain is possible by optimally exchanging information. We then derive analytical expressions to quantify the rate penalty, and analyze how rate penalty changes with a priori knowledge of correlation. In addition, we present a model-based estimation method which can achieve more accurate estimation results compared to directly inspecting the data.

1. INTRODUCTION

Distributed source coding (DSC) addresses the problem of compression of correlated sources that are not co-located. The Slepian-Wolf theorem [1] states that two correlated sources can be optimally encoded (compressed at a rate approaching their joint entropy) even if the encoders only have accesses to the two sources separately, as long as both encoded streams are available at the decoder. Practical code constructions exploiting the Slepian-Wolf theorem have been proposed recently based on channel coding [2, 3, 4], and capacity-approaching code constructions have been reported using turbo codes or low-density parity-check codes (LDPC). These works require *exact* knowledge of correlation available at the encoder, since the correlation information is necessary in setting up the channel code rate. However, in many applications exact correlation information may not be available beforehand, and one would need to estimate it as part of the coding process.[1] While estimation error is inevitable, the impact of estimation error on compression efficiency has not been sufficiently studied for the DSC problem. Note that in many DSC applications only very *limited information exchange* between sources is feasible or desirable (e.g. camera sensors, DSC-based hyperspectral imagery compression [6]), for complexity and power consumption reasons. This means that correlation estimation has to operate under rate constraints.

In this work we derive the optimal information exchange strategy in correlation estimation, and study the rate penalty caused by estimation error. We first derive the optimal strategy for information exchange that minimizes the rate penalty due to inaccurate estimation, under constraints on the total amount of information that can be exchanged between sources. In [5], a similar idea has been investigated to estimate blocking probability in a network. This work proposes a method to allocate sample points to each network link to minimize variance of the estimator. In comparison, our independently derived results are specific to DSC and focus on minimizing an encoding rate penalty. Using the optimal information exchange strategy, we then analyze the relationship between the coding rate penalty and the amount and type of information that was exchanged. We derive analytical expressions to quantify the rate penalty, and analyze how rate penalty changes with a priori knowledge of correlation. In addition, since in many applications we may have information about the statistical models of the data, we present a method that takes advantage of this knowledge, leading to better estimation accuracy as compared to directly inspecting the samples.

This paper is organized as follows. In Section 2 we define the problem and present the estimation process. In Section 3 we derive the optimal information exchange strategy and study the rate penalty due to estimation error. In Section 4 we present a model-based estimation method. Finally, Section 5 concludes the work.

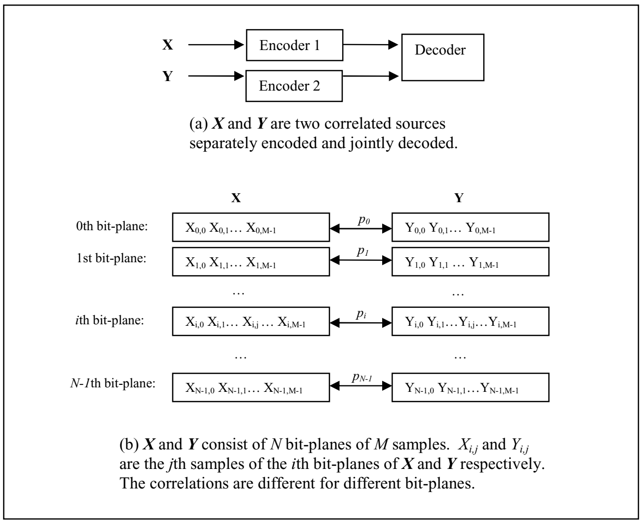
2. PROBLEM DEFINITIONS AND CORRELATION ESTIMATION

Consider the system in Fig. 1, where two correlated sources X and Y are encoded separately and decoded jointly. Here we assume that X and Y are vector sources of size M, which will be coded bit-plane by bit-plane. We will have N bit-planes for each source, with M binary samples in each bit-plane. For the purpose of DSC we are interested in the correlation between bit-planes of same significance in X and Y. Specifically, following the notation in Figure 1, we assume $X_{i,j}, Y_{i,j}$ are i.i.d. equiprobable binary random variables. In addition, $X_{i,j}$ and $Y_{i,j}$ are correlated with crossover probability $Pr[Y_{i,j} = 1|X_{i,j} = 0] = Pr[Y_{i,j} = 0|X_{i,j} = 1] = p_i < 0.5$, i.e. $Pr[X_{i,j} \neq Y_{i,j}] = p_i$.

The crossover probability p_i varies from bit-plane to bit-plane. This situation can arise in many source coding problems. For example, in [6], a DSC based hyperspectral image compression system is proposed by applying set-partitioning on wavelet trans-

[1]Note that in some cases, lack of an accurate correlation model is acceptable if there exists feedback from decoders to encoders [3], but this leads to an increase in overall delay.

0-7803-9134-9/05/$20.00 ©2005 IEEE



(a) **X** and **Y** are two correlated sources separately encoded and jointly decoded.

(b) **X** and **Y** consist of *N* bit-planes of *M* samples. $X_{i,j}$ and $Y_{i,j}$ are the *j*th samples of the *i*th bit-planes of **X** and **Y** respectively. The correlations are different for different bit-planes.

Fig. 1. X and Y are two correlated sources separately encoded and jointly decoded.

formed data to extract bit-planes. In this case $X_{i,j}, Y_{i,j}$ are the samples of the ith bit-planes of the wavelet coefficients of two spectral bands, and different bit-planes are correlated to different extents. Our formulation could be applied to other scenarios where input sources are mapped into a bit-plane representation and correlations between corresponding bit-planes of two sources are exploited via DSC [7, 8].

X and Y are separately encoded and jointly decoded. We compress Y losslessly, which requires $H(Y_{i,j}) = 1$ bits. According to [1], the theoretical limit of encoding X (in bits per sample) is $H(X_{i,j}|Y_{i,j}) = H(p_i) = -p_i log_2(p_i) - (1-p_i) log_2(1-p_i)$. Previous works have reported code constructions that can approach this theoretical limit, when the p_i's are known *exactly* at the encoder. However, in many situations the p_i's are not known exactly and have to be estimated. When there exists an estimation error $\triangle p_i > 0$ there will be a penalty in compression efficiency. On average, this penalty, in bits/sample, is given by:

$$\triangle H = \frac{1}{N} \sum_{i=0}^{N-1} (H(p_i + \triangle p_i) - H(p_i)) \quad (1)$$

Suppose the encoders use the following procedure for correlation estimation: the first encoder samples $n_i \ll M$ samples of $Y_{i,j}$ and sends to the second encoder, which can use this information to encode X. The total number of binary samples exchanged is limited to be n_T, i.e. $\sum_{i=0}^{N-1} n_i = n_T$. We would like to have $n_T \ll N \times M$ in order to keep the information exchange cost small, because this cost is usually non-trivial for DSC applications. For example, in some sensor network applications, the power consumption in inter-node communication is an order of magnitude larger than that of computation. Other examples arise in image/video compression in embedded environments, e.g., hyperspectral image compression in satellites [6], video encoding in mobile devices, etc. In these applications the encoder system may only have enough internal memory to accommodate the data of the current spectral band/video frame (since the application programs and operating systems may have occupied significant portions of the internal memory). In order to estimate the bit-plane crossover probabilities, the system would need to fetch the data of the neighboring frames stored in external memory. Such external memory accesses usually come at the cost of additional power consumption and delay. For example, while some sophisticated CPU/DSPs can handle multiple arithmetic operations in a single cycle, accessing external memory data may incur latency in an order of tens of cycles [9, 10]. So it is desirable to limit the total amount of data exchanged.

By inspecting the n_i pairs $(X_{i,j}, Y_{i,j})$ now available, an estimate of p_i can be computed before X is encoded. We use the upper bound of the $(1-\omega) \times 100\%$ confidence interval as an estimator for a population proportion [11], given by

$$\begin{aligned} \hat{p_i} &= \frac{s_i}{n_i} + z_{\omega/2}\sqrt{p_i(1-p_i)/n_i} \\ &\approx \frac{s_i}{n_i} + z_{\omega/2}\sqrt{\frac{s_i}{n_i}(1-\frac{s_i}{n_i})/n_i} \quad (2) \end{aligned}$$

Here s_i is the number of inspected samples such that $X_{i,j} \neq Y_{i,j}$, and $z_{\omega/2}$ is a constant that depends on the chosen confidence interval, e.g., $z_{\omega/2} = 1.96$ for a 95% confidence interval. Note that we choose the upper bound as the estimator to minimize the risk of decoding failure, at the expense of some encoding rate penalty.

With this estimation, we are $(1-\omega) \times 100\%$ confident (statistically) that the true p_i are within $\frac{s_i}{n_i} \pm z_{\omega/2}\sqrt{p_i(1-p_i)/n_i}$. Hence the estimation error $\triangle p_i = \hat{p_i} - p_i$ is bounded by $0 \leq \triangle p_i \leq 2z_{\omega/2}\sqrt{p_i(1-p_i)/n_i}$ with probability $1-\omega$. In the following we assume

$$\triangle p_i = k\sqrt{p_i(1-p_i)/n_i} \quad (3)$$

where k is a constant that depends on the desired confidence interval.

Note that $\triangle H$, the rate penalty caused by correlation estimation, is a function of (i) $\{p_i\}$, correlation of different bit-planes, (ii) n_T, total number of samples used to estimate correlation, and (iii) $\{n_i\}$, allocation of samples to different bit-planes. In the following section, we investigate: (i) an optimal information exchange strategy, i.e., given $\{p_i\}$, n_T, we derive the optimal $\{n_i\}$ to minimize $\triangle H$; (ii) given the optimal information exchange strategy, we study how $\triangle H$ changes with n_T. Note that the result in (i) requires the knowledge of $\{p_i\}$. In practice, $\{p_i\}$ is obviously unknown. However it is very likely that some of the information known a priori can be used to select adequate $\{n_i\}$'s. For example, the relative values of $\{p_i\}$ may be known (e.g., $p_i < p_j$ if p_i corresponds to a more significant bit-plane than p_j). Moreover, the range of values to be expected may also be known for each i. In the following section we also analyze how sensitive $\triangle H$ is to uncertainty in $\{p_i\}$.

3. OPTIMAL INFORMATION EXCHANGE

3.1. Optimal Information Exchange Strategy

Our first problem is to find the optimal number of samples to exchange, $\{n_i^*\}$, which minimizes $\triangle H$. From (1), we can approximate $\triangle H$ by using Taylor series expansion

$$\triangle H \approx \frac{1}{N} \sum_{i=0}^{N-1} H'(p_i) \triangle p_i, \quad (4)$$

where $\triangle p_i$ is given by (3) and differentiating $H(p_i)$ gives $H'(p_i) = ln(\frac{1}{p_i} - 1)$. To find $\{n_i^*\}$, we solve the following constrained optimization problem: $\min_{\{n_i : \sum_{i=0}^{N-1} n_i = n_T\}} \triangle H$.

By the Lagrangian optimization method, $\{n_i^*\}$ is derived as

$$n_i^* = n_T \frac{\alpha_i^{\frac{2}{3}}}{\sum_{i=0}^{N-1} \alpha_i^{\frac{2}{3}}} \quad (5)$$

where

$$\alpha_i = ln(\frac{1}{p_i} - 1)k\sqrt{p_i(1-p_i)}. \quad (6)$$

Equations (5) and (6) give the optimal sample allocation to minimize $\triangle H$ for a given $\{p_i\}$. It is useful to evaluate how much degradation can be caused by improperly allocating the samples instead of using the optimal allocation. For example, a simple strategy would be to allocate the same number of samples to all bit-planes, i.e., $n_i = n/N$. We define the relative degradation, D, with respect to this uniform sampling by

$$D = \frac{\triangle H_{even} - \triangle H_{optimal}}{\triangle H_{optimal}} \times 100\%$$

where $\triangle H_{even}, \triangle H_{optimal}$ are the $\triangle H$ resulting from evenly and optimally allocating the samples respectively. Experimental results show that evenly allocating the samples can incur significant degradation. For example, for a particular $\{p_i\}$ ={0.475, 0.47, 0.083, 0.08} and with $n_T = 4096$ using 90% confidence interval, D =26.7544%. Note that D would vary according to how the bit-planes are correlated.

Experimental results on real data show similar findings. We have evaluated the relative degradation using a real hyperspectral image compression system [6]. Our preliminary experimental results for this real data set show that the performance degradation matches what is predicted by our model (see [6] for details).

3.2. Rate Penalty Analysis

Now we study how $\triangle H$ changes with n_T. Having an expression for $\triangle H$ as a function of n_T allows the encoder to select appropriate values for n_T, given that increasing n_T leads to additional overhead but also reduces the rate increase due to inaccurate estimation. We can evaluate how $\triangle H$ changes with n_T using the optimal information exchange strategy. This can be determined exactly by (1). Moreover, when $\triangle p_i$ is sufficiently small, $\triangle H$ can be approximated by (4) with $\{n_i^*\}$ given by (5), resulting in:

$$\triangle H \approx \beta / \sqrt{n_T} \quad (7)$$

where

$$\beta = \frac{1}{N} (\sum_{i=0}^{N-1} \alpha_i^{2/3})^{3/2}. \quad (8)$$

Since α_i depends only on p_i, β is independent of n_T. So $\triangle H$ is inversely proportional to $\sqrt{n_T}$.

3.3. Sensitivity Analysis

The optimal sample allocation in (5) requires the knowledge of $\{p_i\}$. However, in practice, a priori knowledge of $\{p_i\}$ may not match the true correlation. Here we analyze how this uncertainty affects the rate penalty. Let $\mathbf{p} = \{p_i\}$ be the vector of "true" correlation, and let $\mathbf{p} + \triangle \mathbf{q} = \{p_i + \triangle q_i\}$ be our a priori estimate of correlation, where $\triangle \mathbf{q}$ is the error. Given $\mathbf{p} + \triangle \mathbf{q}$, we can use $\mathbf{p} + \triangle \mathbf{q}$ in (5), to compute a (sub-optimal) sample allocation $n_i(\mathbf{p} + \triangle \mathbf{q})$. In order to estimate the difference with respect to the optimal allocation, $n_i(\mathbf{p})$, we use a Taylor series expansion

$$\triangle n_i = n_i(\mathbf{p} + \triangle \mathbf{q}) - n_i(\mathbf{p}) \approx \bigtriangledown n_i(\mathbf{p})' \triangle \mathbf{q}. \quad (9)$$

Denote the rate penalty function $\triangle H = f(\mathbf{n}; \mathbf{p}, n_T)$, where $\mathbf{n} = \{n_i\}$ is the vector of sample allocations. The increase in $\triangle H$ due to sub-optimal sample allocation can be approximated by

$$\begin{aligned} \triangle f &= f(\mathbf{n}^* + \triangle \mathbf{n}; \mathbf{p}, n_T) - f(\mathbf{n}^*; \mathbf{p}, n_T) \\ &\approx \frac{1}{2} \triangle \mathbf{n}' \bigtriangledown^2 f(\mathbf{n}^*; \mathbf{p}, n_T) \triangle \mathbf{n} \end{aligned} \quad (10)$$

where $\triangle \mathbf{n} = \{\triangle n_i\}$ is given by (9). (10) can be derived by Taylor series expansion and noticing that $\bigtriangledown f(\mathbf{n}^*; \mathbf{p}, n_T)' \triangle \mathbf{n}$ is zero since $\triangle \mathbf{n}$ is along the direction of linear constraint $\sum n_i = n_T$. Evaluating (10) we obtain

$$\triangle f \approx \frac{\gamma}{n_T^{5/2}} \parallel \triangle \mathbf{n_S} \parallel^2 \quad (11)$$

where $\gamma = \frac{3}{8N}(\sum \alpha_i^{2/3})^{5/2}$ depends on $\mathbf{p}$ only and hence is a constant with respect to a particular allocation, and $\triangle \mathbf{n_S} = [\frac{\triangle n_i}{\alpha_i^{1/3}}]$ is a weighted version of $\triangle \mathbf{n}$. Using (9) and (11) we can evaluate relative degradation D due to error in a priori knowledge, $\triangle \mathbf{q}$. With the previous example $\mathbf{p} = [0.475, 0.47, 0.083, 0.08]'$ and a 5% error in a priori knowledge, $D = 1.6766\%$. Note that using $\mathbf{p} + \triangle \mathbf{q}$ directly to set up channel coding rate may cause decoding error since $p_i + \triangle q_i$ may be less that p_i. Instead, by using $\mathbf{p} + \triangle \mathbf{q}$ to determine a sample allocation and (2) as the estimator to set up channel coding rate we are guaranteed that $\hat{p_i}$ is larger than p_i with probability $(1 - \omega/2)$, and we can bound decoding error systematically.

4. MODEL-BASED ESTIMATION

In many applications we may have information about the statistical models of the data. For example, it is well known that DCT coefficients are well modeled by a Laplacian distribution [12]. In this section we present a method to take advantage of any such a priori model knowledge. The method can result in more accurate estimation than the direct estimation method presented in section 2.

The basic idea is to estimate first the probability density functions (pdf) of the data (X, Y, and $Z = Y - X$), and then use the estimated pdf to derive the crossover probabilities for each bit-plane. Assume that $Y = X + Z$, with X and Z independent. We start by estimating the pdf's $f_X(x)$ and $f_Z(z)$. This can be done by choosing appropriate models for the data samples, and estimating the model parameters using one of the standard parameter estimation techniques, e.g., maximum likelihood estimation (MLE), expectation-maximization (EM), etc.

Once we have estimated $f_X(x)$ and $f_Z(z)$ we can derive the crossover probabilities at each bit-plane as follows. The event that crossover does not occur corresponds to the shaded regions in Fig. 2. Hence we can estimate the crossover probability at bit-plane l by $\hat{p_l} = 1 - I(l)$, where $I(l)$ is given by

$$\begin{aligned} I(l) &= \sum_i \int \int_{A_i} f_{XY}(x, y) dx dy \\ &= \sum_i \int \int_{A_i} f_X(x) f_{Y|X}(y|x) dx dy \end{aligned} \quad (12)$$

The conditional pdf $f_{Y|X}(y|x)$ can be found to be equal to

$$f_{Y|X}(y|x) = f_Z(y - x) \tag{13}$$

and the integral in (12) can be readily evaluated for a variety of densities. In practice we only need to sum over a few regions, A_i, where the integrals are non-zero. Note that the formulation agrees with the observation that when l is small (i.e., least significant bit-planes) the crossover probability is close to 0.5, since in such cases A_i are small and evenly distributed throughout the sample space, and hence for most joint pdf (12) will give $I(l)$ close to 0.5.

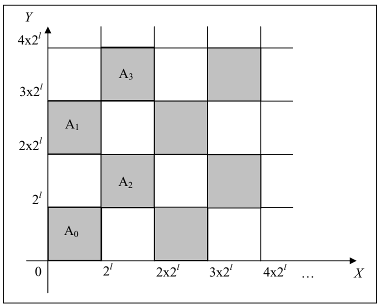


Fig. 2. Crossover probability estimation. The shaded square regions A_i correspond to the event that crossover does not occur at lth bit-plane. E.g., consider $l = 2$ (i.e., the 2nd bit-plane), when X takes the value 0, crossover does not occur when Y takes the value in 0 to 3, or 8 to 11, ..., i.e., when Y is in $m \times 2^l$ to $(m+1) \times 2^l - 1$, where m is an even number.

To compare the model-based estimation with the direct estimation, we generate i.i.d. Laplacian random samples X and Z with different model parameters. We use the crossover probability definition according to [8], i.e., crossover occurs at lth bit-plane when X and Y do not fall into the same quantization bin of size 2^l. In the model-based estimation MLE is used to estimate the model parameters. Fig. 3 compares the estimation results at the 5th bit-plane. The deviation here is with respect to the empirical crossover probability calculated using all the samples. As shown in the figure, substantial reduction in the deviation is possible using the model-based estimation. This is because at the 5th bit-plane the crossover probability is of the order of 10^{-3}. Thus direct estimation would need to exchange several thousands samples in order to obtain reliable estimation results. In general model-based estimation can achieve significant improvement when the crossover probability is small. Those are also the situations when we can obtain significant data compression using DSC.

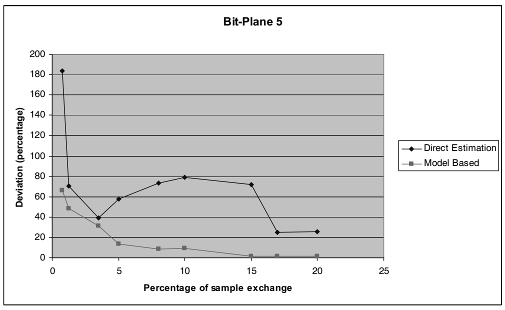


Fig. 3. Comparing direct estimation and model-based estimation.

5. CONCLUSIONS

We have derived the optimal information exchange strategy that minimizes the rate penalty due to inaccurate estimation, under constraints on the number of samples that can be exchanged between sources. Experimental results have shown that, for a particular correlation, an arbitrary sample allocation can cause significant degradation (e.g., 26% increase in rate penalty). Experimental results on hyperspectral image data show similar findings. We have also presented analytical expressions to quantify the rate penalty due to estimation error. Rate penalty is approximately inversely proportional to the square root of the total number of sample exchanged. Note that the result in optimal information exchange requires the knowledge of correlation, so we have also analyzed how a priori correlation knowledge affects the rate penalty. In addition, we have presented a model-based estimation method that can reduce estimation error significantly compared to directly inspecting the samples. The model-based estimation is particularly useful in image/video compression, where knowledge of the statistical model of the data is often available.

6. REFERENCES

[1] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471-480, July 1973.

[2] S. Pradhan and K. Ramchandran, “Distributed source coding using syndromes (DISCUS): Design and construction,” in *Proc. DCC*, Mar. 1999, pp. 158-167.

[3] A. Aaron and B. Girod, “Compression with side information using turbo cdodes,” in *Proc. DCC*, Apr. 2002, pp. 252-261.

[4] A. Liveris, Z. Xiong and C. N. Georghiades, “Compression of binary soruces with side information at the decoder using LDPC codes,” *IEEE Communication Letters*, vol. 6, Oct 2002,pp. 440-442.

[5] P. Lassila, J. Karvo and J. Virtamo, “Efficient importance sampling for Monte Carlo simulation of multicast networks,” in *Proc. IEEE INFOCOM 2001*, Apr. 2001, pp. 432-439.

[6] C. Tang, N.-M. Cheung, A. Ortega and C. S. Raghavendra, “Efficient inter-band prediction and wavelet based compression for hyperspectral imagery: A distributed source coding approach, ” in *Proc. DCC*, 2005.

[7] A. Sehgal, A. Jagmohan, and N. Ahuja, “Wyner-Ziv coding of video: an error-resilient compression framework,” in *IEEE Trans. Multimedia*, 2003.

[8] H. Wang, N.-M. Cheung and A. Ortega, “WZS: Wyner-Ziv scalable predictive video coding,” in *Proc. PCS*, 2004.

[9] N. T. Slingerland and A. J. Smith, “Cache performance for multimedia applications,” in *Proc. IEEE International Conference on Supercomputing*, 2001.

[10] *Hardware Analysis*. Available at http://www.hardwareanalysis.com/content/reviews/article/1499.5/

[11] W. Mendenhall and T. Sincich, *Statistics for Engineering and the Sciences*. Prentice-Hall, 1995.

[12] E. Y. Lam and J. W. Goodman, “A mathematical analysis of the DCT coefficient distributions for images, ” in *IEEE Trans. Image Processing*, 2000.